Nucleic Acids Research, 1996, Vol. 24, No. 17 **3295–3301**

A revised secondary structure model for the 3′-end of hepatitis B virus pregenomic RNA

Alistair H. Kidd[1] and Karin Kidd-Ljunggren[1,2,*]

[1]Department of Virology, University of Umeå, 90 185 Umeå, Sweden and [2]Department of Infectious Diseases, University of Lund, 221 85 Lund, Sweden

Received June 11, 1996; Revised and Accepted July 11, 1996

ABSTRACT

The polymerase encoded by human hepatitis B virus, which has reverse transcriptase and RNase H activity, binds to its pregenomic RNA template in a two-step process involving a terminal redundancy. Both first strand and second strand DNA synthesis involve primer translocation and second strand synthesis involves a template jump. Three parts of the genome, including the so-called core promoter, are known to show deletions in strains usually arising after long-standing HBV infection, but also in some patients treated with interferon. A computer-based study of RNA template folding in the core promoter region, accommodating well-known point mutations, has generated a model for the 3′ DR1 primer binding site as being part of a superstructure encompassing an already well-established stem–loop. Depending on the identity of nucleotides 1762 and 1764, the DR1 region may assume two alternative secondary structures which stabilize it as a primer binding site to different extents. Remarkably, one of these structures includes a pronounced loop which coincides with at least 12 related deletions seen in HBV DNA from different patients. Thus according to the model, the 5′- and 3′-ends of pregenomic RNA, which share primary sequences but have separate functions, are not structural equivalents. An RNA superstructure near the 3′-end of all HBV transcripts could have far-reaching implications for the modulation of both genome replication and post-transcriptional processing.

INTRODUCTION

Hepadnavirus genomes have a compact organization in which all transcriptional regulatory signals coincide with open reading frames (ORFs). Almost half of the 3.2 kb partially double-stranded DNA genome of human hepatitis B virus (HBV) has overlapping ORFs. The ORFs are termed precore/core, polymerase, pre-S/S and X. There are five major unidirectional transcripts, including the 3.5 kb pregenome and a 3.5 kb RNA species termed precore RNA, which is 30 nucleotides (nt) longer (1). In common with precore RNA, pregenomic RNA acts as a messenger RNA, but it functions additionally as a template for reverse transcription (Fig. 1). This occurs in immature core particles and encapsidation of the hepadnavirus pregenome and polymerase together is required for DNA synthesis (2).

The pregenomic RNA of hepadnaviruses carries a *cis*-acting encapsidation signal near its 5′-end (3,4). The polymerase is also required for packaging of pregenomic RNA (5) and a complex between the polymerase and the encapsidation signal forms before encapsidation occurs (2). Formation of this ribonucleoprotein complex involves cellular proteins, including the heat shock protein Hsp90 (6).

The encapsidation signal has inverted repeat sequences which form a bipartite stem–loop structure (7,8). This structure appears to be well conserved in different hepadnaviruses and highly conserved in naturally occurring HBV variants (9). By virtue of a terminal redundancy in pregenomic RNA, the stem–loop structure also occurs at the 3′-end. However, it is the 5′ copy which functions in polymerase binding and encapsidation (10). To start reverse transcription, a 4 nt primer which is covalently attached to the polymerase is made, using a bulge in the encapsidation signal as template (Fig. 1). This primer is then shifted to DR1, near to the 3′-end of the RNA, and reverse transcription proceeds (2). The mechanism by which the 4 nt primer is translocated between the 5′- and 3′-ends is not known, but may involve circularization of the pregenome.

Once minus strand synthesis is complete, most of the pregenomic template has been removed by RNase H activity of the polymerase, except for its 5′ DR1 region. This remaining RNA can act as primer for second (plus) strand synthesis, but a shift in the primer position occurs, from DR1 (3′) on the minus strand template to DR2, a direct sequence repeat closer to the 5′-end (Fig. 1). Synthesis of plus strand DNA in a continuous fashion can then only be achieved by a jump from the 5′- to the 3′-end of the minus strand template. While this is mechanistically complex, it ensures that single-strand breaks in minus strand and plus strand DNA do not coincide, so that non-covalently closed DNA duplexes have circular form.

Since the advent of the PCR technique, many workers have focused attention on the genetic variability of HBV. HBV strains fall naturally into five or six phylogenetic groups or genotypes (A–F) (11–13). Strains of the main genotypes (A–D) are generally associated with different geographical areas of the world (14).

Infection by HBV has a wide range of clinical outcomes, from self-limited silent or acute infection to fulminant hepatitis. It has been estimated that >300 million cases of chronic HBV infection exist globally. During attempts to explain viral factors which might lead to severe infection or fatality, mutations have been

* To whom correspondence should be addressed at: Department of Virology, University of Umeå, 901 85 Umeå, Sweden

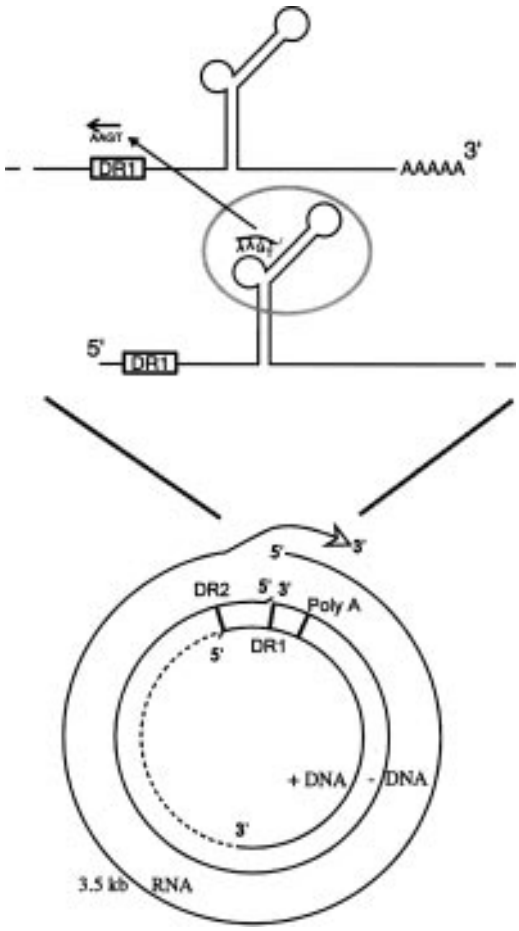

Figure 1. Organization of the terminally redundant 3.5 kb pregenomic RNA relative to the 3.2 kb HBV genome and key events for transcription, encapsidation and reverse transcription. Pregenomic RNA transcription from circular DNA (lower part of diagram) starts immediately upstream of DR1 and continues around the genome through DR1 again to beyond the polyadenylation signal. The polyadenylation signal at the 5′-end of the RNA is not recognized. The viral polymerase, shown with oval shading in the expanded (upper) part of the diagram, associates with a stem–loop structure at the 5′-end of the pregenomic RNA and is encapsidated with it. A 4 nt primer (5′-TGAA-3′) covalently linked to the polymerase is generated using part of the encapsidation signal as template. This primer is moved to a corresponding sequence (part of DR1) near the 3′-end of the RNA and minus strand DNA synthesis begins.

discovered which apparently arise during the course of infection and are sometimes associated with changes in immune response to the virus. The precore codon 28 mutation (15) is a single base change predicted to prevent e-antigen (HBeAg) production (see Table 1) and is usually associated with loss of serum HBeAg and a concomitant increase in antibody production (anti-HBe) by the patient. Another DNA change ($A^{1762}GG^{1764}$→TGA or TGG) in the core promoter region is also associated with HBeAg to anti-HBe seroconversion (16). In addition, several workers have described deletions arising in the pre-S/S region of the genome, the core region of the precore/core ORF (17) and also around the $A^{1762}GG^{1764}$ region of the core promoter (16,18,19).

Despite many publications on HBV genome variability, few satisfactory explanations have been put forward for the generation and tolerance of common point mutations and deletions arising during the course of a single infection. The precore codon 28 stop mutation, which coincides with the main stem of the encapsidation signal, has been investigated most thoroughly. Lok *et al.* (20) carried out a large scale survey to determine the frequency of mutations in this part of the genome. The co-existence of certain mutations in non-contiguous codons (precore codon 15 and codons 28 and 29) was related to the secondary structure of the pregenomic encapsidation signal. In all strains except those of genotype A, U-A pairing would normally replace U-G pairing, which in addition to preventing HBe production at the translational level, might improve the encapsidation competence of the mutant (21).

Table 1. Open reading frames in the HBV genome

ORF	Nucleotide position[a]		Product
	Start	End	
X	1374	1838	X protein
Precore	1814 →[b]		HBeAg[c]
Core	1901	2452	Core (HBcAg), HBeAg[c]
Pre-S1	2848 →		L protein
Pre-S2	3205 →		M protein
S	155	835	S protein (HBsAg)
P	2307	1623	Polymerase

[a]Nucleotide numbering according to Okamoto *et al.* (1988).
[b]Arrows indicate ORFs that are contiguous with and use the stop codon of the next on the list without an arrow.
[c]HBeAg is a C-truncated core product, originating at the precore start codon but post-translationally cleaved to carry only the last 10 amino acids of 29 encoded by the precore.

The core promoter mutation $T^{1762}GA^{1764}$ from AGG and deletions around this region have been assumed to be important at the transcriptional level for modulation of precore and/or pregenomic RNA levels (16). However, Chen *et al.* (22) recently defined the essential promoter elements as coinciding with the start of the precore RNA, some 24 bases downstream of the point mutations. In this paper we describe an alternative role for the AGG/TGA motif during reverse transcription of the pregenomic RNA. A model based on computer prediction is proposed for the secondary structure of the 3′-end of this RNA, which explains the TGA mutation and why deletions may subsequently arise in this region. The model also offers a plausible alternative explanation for the dramatic decrease in virus replication of genotype A strains on developing a precore codon 28 stop mutation (21).

MATERIALS AND METHODS

HBV strains and sequence determination

Several HBV strains used in this study have been described previously (14). As the X gene/core promoter sequences were already established, sequencing was extended downstream to include the precore region/encapsidation signal. All sequencing was performed by the method of Kretz *et al.* (23) as described previously (24).

Nucleotide numbering throughout corresponds to that of Okamoto *et al.* (11). This differs by 2 nt from the numbering system adopted by Chen *et al.* (22) for core promoter analysis.

Phylogenetic analysis

HBV X gene sequences were aligned with those of several known strains (14) using CLUSTALV (25) and subsequently analysed

using the PHYLIP package, version 3.5 (26). The programs SEQBOOT, DNAPARS, CONSENSE and DRAWTREE were used consecutively as described previously (14).

Computer prediction of RNA secondary structure

Computer models of RNA folding were generated using the MFOLD program, a part of the GCG package (27), which is based on the revised predictions of Jaeger and Zuker (28). This program allows consideration of suboptimal folding of RNA. The on-line facility of M. Zuker (Institute for Biomedical Computing, Washington University, St Louis, MO; http://www.ibc.wustl.edu/~zuker/RNA/form1.cgi) was used for the purpose of confirmation, as was RNADRAW (29), a graphical interactive program based on the Vienna RNA package. All calculations were performed for a temperature of 37°C.

RESULTS

Taking into account that AGG→TGA changes at positions 1762–1764 in the so-called core promoter are so common and that many natural deletions arise across this same region (nt 1748–1777), we undertook a computer-based study of secondary structure prediction. Our main premise was that there must be a unifying structural explanation for these two apparently linked phenomena which is compatible with our present knowledge of HBV genome transcription and replication. Since the initiator and TATA elements of precore and pregenomic RNA transcription are now well defined (22), the replication events following transcription and mediated by the viral polymerase might rather be the reason for such mutations arising.

Seventy-five HBV sequences (including 45 of our own) spanning the core promoter, the precore region and the first 28 nt of the core region were aligned and compared. These sequences represented genotypes A–D, as determined by phylogenetic analysis (14). From this alignment, a map of genotype-specific changes was made (Fig. 2). Much of the computer-based analysis was arbitrarily based on a genotype A HBV sequence from the databank (accession no. V00866). This 3.2 kb genomic sequence was edited to resemble 3.5 kb pregenomic RNA, then subedited for analysis in sections. Most of what follows deals with analysis of a 375 base sequence representing the 3′-end of the pregenomic RNA, from nt 1582 to 1956 according to the genomic numbering (11).

Folding of the 375 base sequence by the program MFOLD (28) was performed using a sequence with (i) AGG or (ii) UGA at positions 1762–1764. The well-established stem–loop structure corresponding to the 5′ encapsidation signal was present in 15 of 16 suggested likely structures for (i) and in all 16 suggested structures for (ii). Surprisingly, 12 of the 16 suggested structures for (i) also had a longer stem–loop structure in common, situated immediately upstream of the established one, from positions 1746 to 1845 (Fig. 3a). Even more surprisingly, according to the most likely predictions, this stem–loop would share a short stem and basal bulge with that of the established stem–loop. Eleven of the 16 predictions for (ii), with a UGA at positions 1762–1764, also featured a novel stem–loop structure. The predicted UGA-specific stem–loop structure had all basal stem features in common with its AGG-specific equivalent (Fig. 3b). However, the predicted folding between positions 1758 and 1833 was remarkably different.

For clarity, the newly predicted stem–loop structures are hereafter referred to as stem–loop 1, whereas the established (downstream) one will be referred to as stem–loop 2. The possible

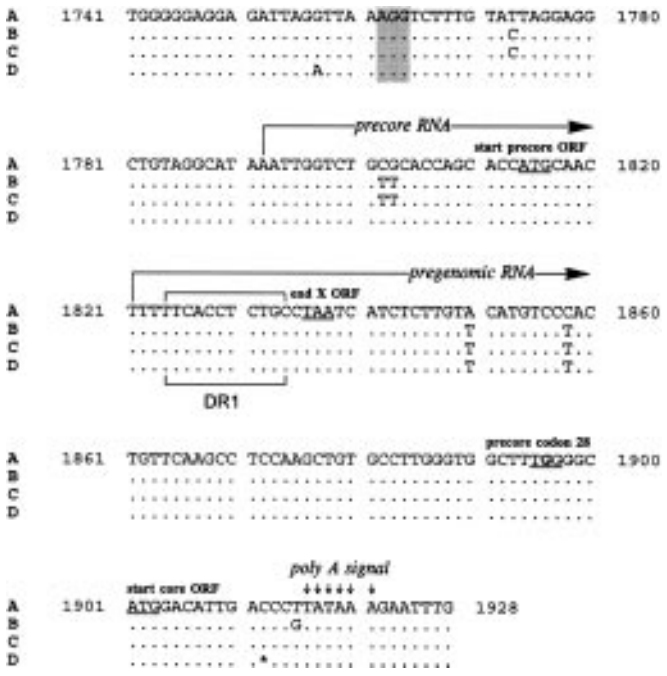

Figure 2. Genotype-specific differences in core promoter/precore sequences. The alignment shows a consensus sequence for each of the main genotypes A ($n = 8$), B ($n = 21$), C ($n = 27$) and D ($n = 19$), based on 45 of our own sequences (from separate patients) and 30 from the databank. Position 1762 is always shown as A here and 1764 as G (shaded region); these are frequently T and A respectively, regardless of genotype. Position 1896 is always shown as G here; an A at this position is frequently encountered in genotype B, C and D strains (precore codon 28 stop mutation). Additionally, position 1912 in genotype D (denoted by an asterisk) can be either C or T.

structures of both stem–loop 1 and stem–loop 2 combined (Fig. 3a and b), as predicted by MFOLD for the 3′-end of pregenomic RNA, will be called a superstructure.

The differences in predicted stem–loop 1 folding as a function of the identity of nt 1762 and 1764 are worth detailed consideration. According to the folding predictions, AGG at positions 1762–1764 would be part of a 4 nt sequence (5′-AGGU-3′) base pairing with the middle of the 11 nt DR1 sequence (5′-UUC**ACCU**CUGC-3′) (Fig. 3a). The two bases at the 3′-end of DR1 would also be capable of base pairing (5′-UUC**ACCU**CU**GC**-3′). The predicted shift in secondary structure as a result of UGA identity at positions 1762–1764 would transfer these bases away from DR1 to a newly created stem–loop side branch (covering 1759–1775). In this second conformation for stem–loop 1, the DR1 sequence is base paired in all positions except for the four 5′-nucleotides which bind the primer (5′-UUCA**CCUCUGC**-3′). The middle region of DR1 (CCU), as before, pairs with an AGG sequence, but $A^{1762}GG^{1764}$ is substituted by $A^{1778}GG^{1780}$.

We then asked whether the predicted folding of the 'UGA' variant of stem–loop 1 was dependent on a change at both nucleotide positions 1762 and 1764 or whether one base change would suffice. The folding predictions indicated that *either* a change from A to U at position 1762 *or* a change from G to A at position 1764 was necessary and sufficient to effect the structural rearrangement. However, a change at 1764 alone would form a more open loop, rather than a side branch, between nt 1759 and

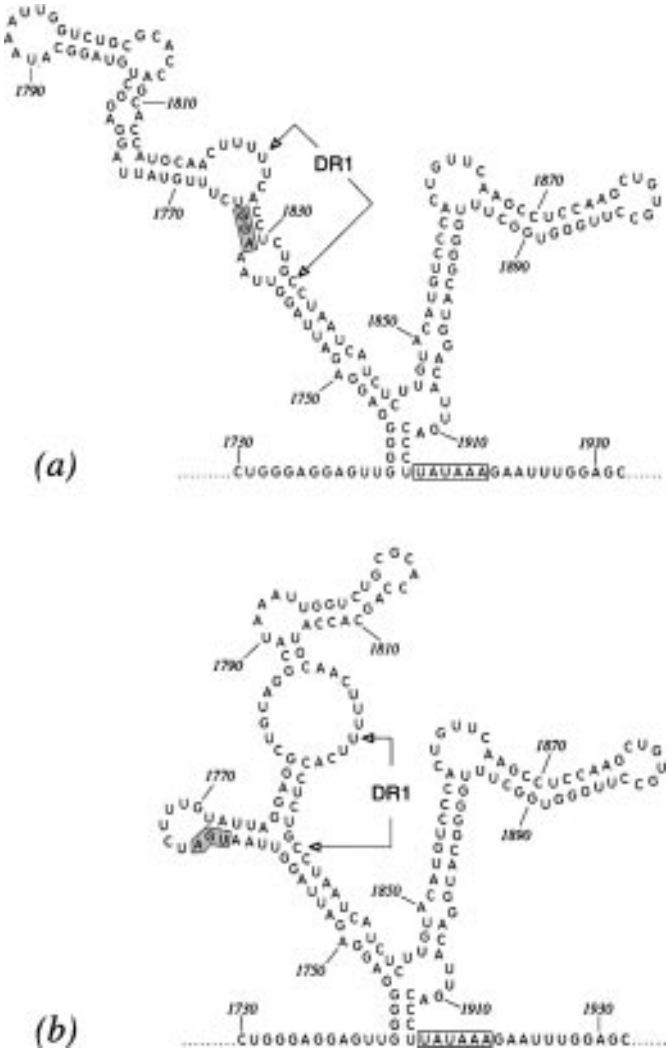

Figure 3. Computer predictions (using MFOLD) of secondary structure for the 3′-end of HBV pregenomic RNA with: (**a**) AGG at positions 1762–1764 (shown shaded); (**b**) UGA at positions 1762–1764 (also shown shaded). All HBV mRNA species carry these primary sequences and use the polyadenylation signal at positions 1916–1921 (shown boxed). The computed free energy for optimal folding from nt 1750 to 1841 inclusive is –18.3 kcal/mol in (a) and –17.1 kcal/mol in (b).

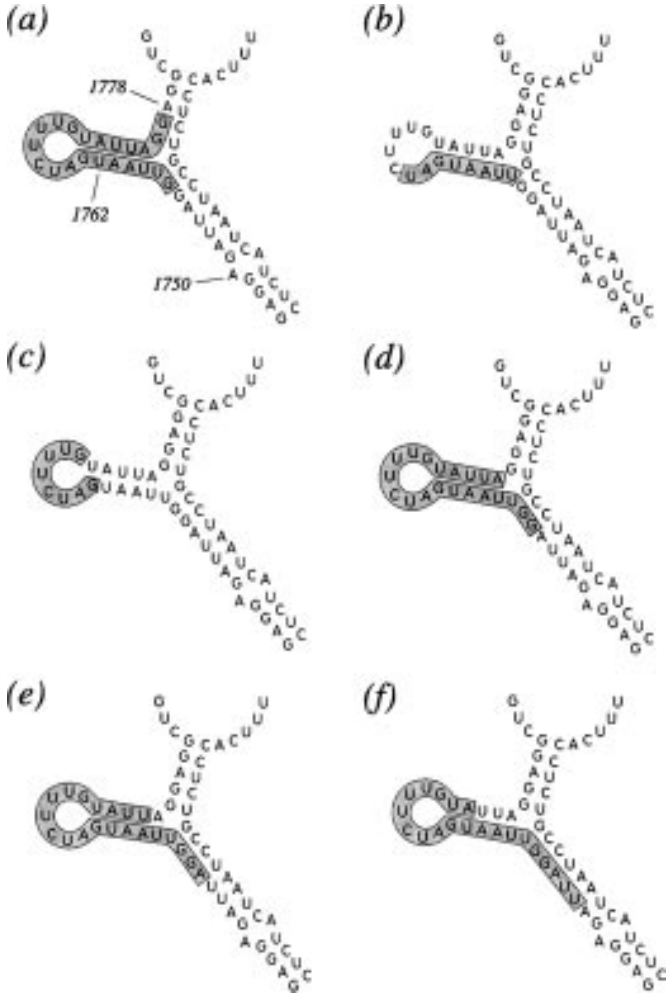

Figure 4. Location of six different representative natural core promoter deletions in the HBV genome relative to part of the computer predicted secondary structure shown in Figure 3b, which corresponds to strains with UGA at positions 1762–1764. References for these deletions are as follows: (**a**) (16); (**b**) (16); (**c**) (16,18); (**d**) (31); (**e**) (19); (**f**) (16). MFOLD analysis of modified HBV V00866 sequences, with the respective deletions, indicated that the uppermost part of the structure shown here is conserved. This involves base pairing between nt 1778–1780 and the central part of DR1. The stem carrying nt 1750 often appeared shortened as a result (data not shown).

1774, as a result of a base mismatch at position 1762 (A^{1762} with A^{1772}) (data not shown). This folding is based on a genotype A sequence, however, and we note that genotype B and C strains generally have C at position 1773 rather than U. The C would not pair with A^{1761}, so there would be a loop irrespective of nucleotide identity at position 1762.

Computer predictions of RNA secondary structure must always be interpreted with caution and backed up by other forms of evidence for their existence. In this case, the pattern of deletion mutations seen in DNA from patients with long-standing infections requires consideration. At least 12 different deletions have been described which cover nt 1748–1777 (16,18,19,30–33). When these deletions were mapped on stem–loop 1, the extra branch associated with UGA at positions 1762–1764 was highlighted. Figure 4 shows the positions of six such deletions relative to the secondary structure prediction for the intact sequence (Fig. 3b). Thus, the deletions occur in a region of the pregenomic RNA which may exist as a side branch or loop. There was no such association between the positions of deleted sequences and the secondary structural features depicted in Figure 3a, which were predicted by MFOLD for AGG at positions 1762–1764.

The concept of a 3′ superstructure thus gains credence from the fact that it is compatible with the emergence of mutations during infection which until now have defied explanation. Deletions in this region may be a natural consequence of template skipping by the polymerase during reverse transcription of a region with a natural tendency to three-dimensional folding. This prompted us to investigate the model further for predicted secondary structure shifts on alteration of base identity at positions known to vary often, either within genotypes or across genotypes, during the course of natural

infection (Fig. 2). A limited computer-based 'multivariance analysis' for RNA folding was performed using the 375 bp V00866 sequence. Positions 1762–1764 could be either AGG or UGA, position 1896 could either be G or A (corresponding to the precore codon 28 stop), position 1858 could either be U or C (the latter being a genotype A- and genotype F-specific change) and position 1899 could be either G or A.

All sequences corresponding to genotypes B, C and D at position 1858 showed a stem–loop 1 configuration of either the 'AGG' or 'UGA' pattern, depending on nucleotide identities at positions 1762–1764 (corresponding to Fig. 3a and b). This result was irrespective of base identity at positions 1896 and 1899. However, the genotype A-specific sequence (corresponding to C at position 1858), while retaining its basic stem–loop 1 configurations with AGG/UGA changes at positions 1762–1764, had a predicted shift in its secondary structure as a result of the precore stop codon mutation (G→A at position 1896). Although keeping its distal secondary structure, stem–loop 2 was predicted to undergo a reconfiguration involving all base pairing in its lower half (Fig. 5a and b). Such a reconfiguration would effectively destroy any common stem organization of the superstructure and alter the local environment of the polyadenylation signal.

In the light of the above results, we re-examined stem–loop folding at the 5′-end of the pregenomic RNA, using 166 bases from the V00866 sequence. Pregenomic RNA starts at nt 1821, some 27 bases upstream of the 5′-end of the well-known stem–loop. Thus, a stem–loop 1 cannot exist at the 5′-end of pregenomic RNA, which precludes the possibility of a superstructure. Contrary to our previous results with 3′ sequences, no gross alterations in the well-known stem–loop structure at the 5′-end could be emulated by the program MFOLD when a genotype A-specific sequence (C at position 1858) was made to mutate at precore codon 28 from G to A (data not shown).

We also examined the 3′-terminal sequences of precore RNA for woodchuck hepatitis virus (WHV) (accession no. J02442) and ground squirrel hepatitis virus (GSHV) (accession no. K02715). Stem–loop regions 1 and 2 could be discerned by computer folding. As for HBV, the 5′-end of DR1 was seen to be part of a bulge. Two consecutive CCU motifs, within DR1 and immediately 3′ of it, were seen to be associated with AGG motifs (Fig. 6), as was the case for HBV (Fig. 3a and b).

Computer folding for the much more distantly related duck hepatitis virus (DHBV) pregenomic RNA (accession no. M32991), which does not encode an X ORF, gave results different to the others. In the latter case, DR1 was predicted to map to an 'open' region with no obvious secondary structure (data not shown).

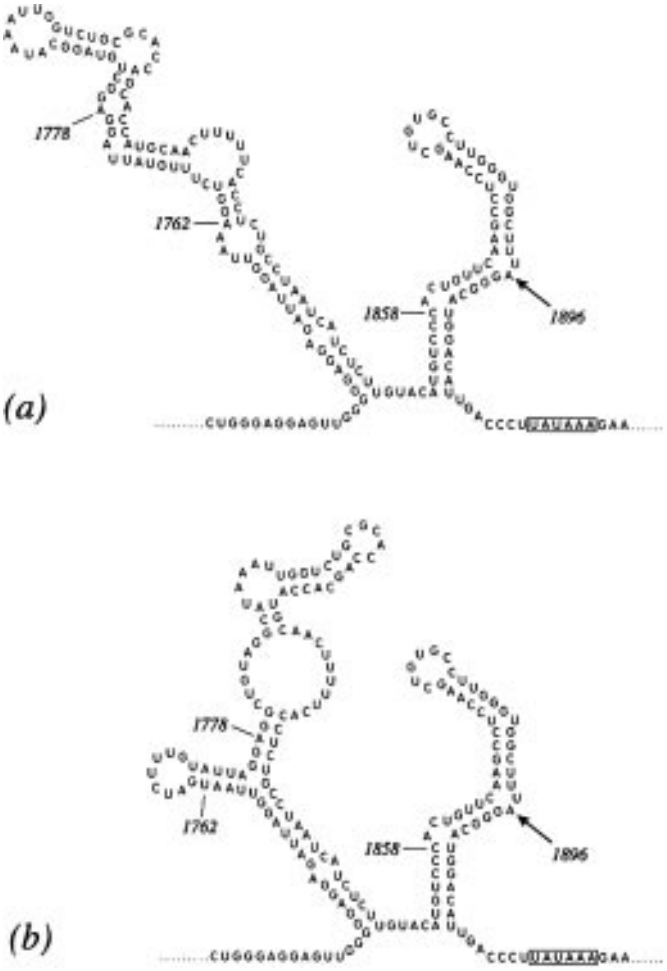

Figure 5. Predicted change in the secondary structure of stem–loop 2 as a result of a single mutation at position 1896 (precore codon 28 stop) and peculiar to genotype A strains. (**a**) Predicted superstructure with nucleotide identity AGG at positions 1762–1764, after altering position 1896 from G to A (shown by arrow; compare with Fig. 3a). (**b**) Predicted superstructure with nucleotide identity UGA at positions 1762–1764, after altering position 1896 from G to A (compare with Fig. 3b). Note the dissociation of stem–loop 1 from stem–loop 2 and the altered position of the polyadenylation signal compared with Figure 3. The reduction in thermal stability between optimal structures (Fig. 3 versus Fig. 5) as a result of the mutation at position 1896 would be 4.3 kcal/mol.

DISCUSSION

In current research on hepadnavirus replication, much interest lies in the elucidation of how the virus polymerase first interacts with the 5′ stem–loop structure of pregenomic RNA and accomplishes a 4 nt primer synthesis. The events of encapsidation are tightly linked to these early events of reverse transcription and are obvious targets for intervention. Many questions about this process remain unanswered. Why, for example, does the polymerase prefer the 5′ copy of the encapsidation signal to the 3′ copy, despite the fact that a 3′ copy can be made to function and cause priming of DR1 in artificial constructs (34)? If the 3′ copy is not required for viral replication, what strategy does the virus use to avoid its use or any competitive effect that it might have? One simple explanation could be that the 3′ redundancy in pregenomic RNA shares the same primary sequence, but not the same three-dimensional conformation.

Related to this first enigma is the additional one of polyadenylation signal usage. The one polyadenylation signal, which is somewhat unusual ($U^{1916}AUAAA^{1921}$), is present at the 3′-end of all HBV transcripts, but both pregenomic RNA and precore RNA carry it at the 5′-end also. The reason why the upstream copy of the signal is ignored was addressed by Russnak and Ganem using WHV (35), who found that proper usage depended on multiple sequences 5′ of it, which increased its efficiency of use. These sequences were located within 400 nt of the polyadenylation signal. The authors suggested a stem–loop (now known to

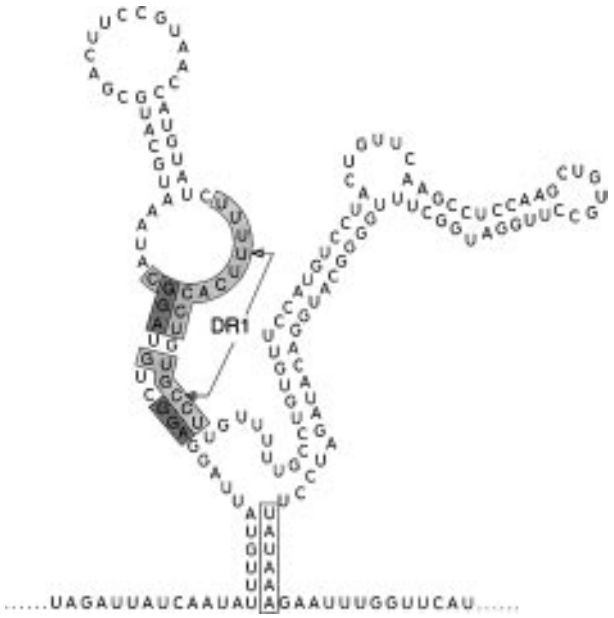

Figure 6. Predicted 3′ secondary structure for WHV pregenomic RNA. Base pairing at the polyadenylation signal (boxed) is predicted for WHV, but not HBV (see Fig. 3a and b). The grey shading indicates nucleotide positions with conserved identity around DR1 between the predicted secondary structures for WHV and GSHV and also for that shown in Figure 3b for HBV. Two AGG motifs stabilizing DR1 (dark shading) are also a feature of the HBV RNA folding depicted in Figure 3a and b.

correspond to the 5′ encapsidation signal) immediately upstream of the poly(A) signal as having a possible role in activating it.

The structures proposed in this paper, invoking the concept of structural differences at the 5′- and 3′-ends of pregenomic RNA, are of relevance in answering the above questions. These structures can be regarded as variants of a working model, which both deserve and demand experimental confirmation. The analysis was originally performed to check whether core promoter mutations, both point mutations and deletions, might arise as a result of secondary structural considerations. The computer work presented here is convincing evidence that they may.

We propose that the creation of deletions in the core promoter region of HBV is a result of template skipping by the polymerase. This is compatible with what we know of its natural function: template jumping is thought to be the mechanism by which plus strand synthesis in hepadnaviruses is continued (2). The proposal is also compatible with the computer generated secondary structures, where natural deletions arising in the core promoter map to a well-defined loop region (Fig. 4). Since completing this first analysis, we have examined the pre-S/S region of HBV for predicted RNA secondary structure and, again, both a 129 base deletion (36) and a completely overlapping 183 base form of the deletion (37–39) mapped to a long predicted stem–loop structure (data not shown).

Since the genome of HBV has limited coding capacity (with overlapping ORFs and regulatory signals), it is reasonable to assume that only certain regions can be deleted and perhaps even only at certain stages of infection or under certain circumstances. Thus, although template skipping may arise with some frequency throughout the pregenomic RNA as a result of secondary structures, such events are probably almost invariably lethal for the continuation of that genome once expressed. It is possible that deletions are only tolerated in the natural history of HBV infection when the need for certain products or parts of products declines. In the case of the so-called core promoter deletions mapping within the X ORF, there could be a decreased need for those functions of the X protein that are associated with its C-terminal domain (40–42). However, deletions could also be of advantage to the survival of HBV. For the pre-S region, deletions are probably related to the removal of epitopes and escape from the immune response (17). Deletions in the HBV genome have been seen to arise during or after treatment with interferon (36). One reason could be that double-stranded RNA, which is present in any RNA with appreciable secondary structure, potently promotes interferon action through more than one pathway (43). Accidental loss of such secondary structures, if not entirely lethal for replication, might thus be of advantage to the survival of HBV during interferon therapy.

By attempting to explain mutations in the HBV core promoter region, we have developed a model for the 3′-end of pregenomic RNA of HBV, WHV and GSHV involving two closely linked stem–loop regions. These stem–loop regions (referred to as 1 and 2 in this paper) may undergo base pairing to form a common stem (Figs 3 and 6), but this may not be a functional requirement. The most important consideration is the environment proposed for DR1 within stem–loop 1 of this secondary structure. This involves conserved base pairing (Fig. 6) and could even involve weak three-dimensional interaction with nucleotides of stem–loop 2. The latter possibility has until now been considered to be unimportant for the replication process, based largely on work with DHBV.

The existence of stem–loop 1 at the 3′-end of pregenomic RNA will be more difficult to test experimentally than the existence of stem–loop 2. Stem-loop 1 as proposed for HBV and WHV does not have the same stringency of base pairing as stem–loop 2 and could be predicted to be more of a dynamic structure. We note that no predicted Watson–Crick base pairing extends over more than seven bases, which may have relevance for easy unfolding of the structure during reverse transcription. The 5′-end of DR1 in our model, like the site of 4 nt primer synthesis, is part of a bulge, which may have importance for the poorly understood mechanism of primer transfer, at least in mammalian hepadnaviruses.

With regard to the well-known point mutations in HBV, it is worth considering why a UGA motif at positions 1762–1764 might be an advantage over an AGG motif, which was predicted to base pair with part of the DR1 sequence (Fig. 3a). The most likely explanation for this would be that a change to UGA causes a secondary structure shift (Fig. 3b) and that the substituting nucleotides can support the whole of the DR1 region by base pairing, except for the four nucleotides directly involved in priming of the minus strand RNA. This would have the advantage of minimizing molecular movement in this region even further and at the same time making all four primer binding nucleotides free from predicted base pairing (5′-UUCA-3′), instead of just three (5′-UUC-3′).

Why, then, would the UGA motif not be favoured in the ‘wild-type’ sequence, rather than arising during infection? The answer to this may be that this region also encodes the X protein and that the AGG/UGA change causes a non-conservative change in two amino acids ($K^{130}V^{131}$) thought to reside in an important domain of the protein for its transactivation function (40–42). Thus, during establishment of infection, it may be an advantage (or a necessity) for the virus to have X protein activity in infected cells, whereas during long-established infection and

especially during antiviral treatment, efficiency of nucleic acid replication may be a more important factor.

It is interesting to note that computer folding predicted that both position 1762 alone and 1764 alone can determine whether stem–loop 1 would assume a long or a branched structure (Fig. 3a and b). The motif AGA at 1762–1764, however, creates a *loop* as a branch, rather than a *stem–loop*. The branch would probably be better stabilized on a further change of the motif to **U**GA. Whether a stem–loop is possible also involves consideration of nucleotide identity at position 1773 (C or U), since the majority of genotype B and C strains have C at this position, which cannot base pair with nt 1761 (A).

To our knowledge, no core promoter deletions have been detected which eradicate the $A^{1778}GG^{1780}$ motif, which we postulate to be instrumental in stabilizing the central part of the DR1 sequence in Figure 3b. Since deletions have been described involving sequences up to and including position 1777 (Fig. 4; 16,18,19,30–33) and since the majority of these deletions have been predicted here by computer folding to preserve the interaction between $A^{1778}GG^{1780}$ and the central motif of DR1 ($C^{1828}CU^{1830}$), motif $A^{1778}GG^{1780}$ may indeed play an indispensable structural role in the conformation depicted in Figure 3b. We note that deletion of sequences bearing $A^{1762}GG^{1764}$ (between positions 1748 and 1777) in the conformation of Figure 3a could lead directly to the conformation of Figure 3b. However, while there is evidence that such deletions arise from strains carrying U^{1762} and/or A^{1764} (16), it is not clear from the literature how often deletions can arise from strains carrying AGG at positions 1762–1764.

As a model, the superstructure with partially base paired DR1 not only explains the existence of specific core promoter point mutations and deletions, but it also offers an additional explanation to that of Li *et al*. (21) and Lok *et al*. (20) for the importance of the precore stop mutation in virus replication. There appears to be a substantial decrease in the replication competence of genotype A HBV strains on developing the codon 28 stop mutation (21). While this mutation in genotype B, C and D strains might, by virtue of creating a stronger U-A base pair from U^{1858}-G^{1896}, stabilize encapsidation in the 5′ copy of stem–loop 2, the C-A pair created in genotype A may have significance at the 3′ copy of the stem–loop. A radical alteration in stem–loop 2 (and destruction of its close association with stem–loop 1) after the so-called precore codon 28 mutation might serve to alter the three-dimensional RNA conformation of DR1 or its immediate environment and therefore its ability to accept the 4 nt primer. Additionally, it is not yet clear whether local secondary structure could have an effect on the functioning of the polyadenylation signal which serves for all HBV mRNAs, but testing of this is clearly necessary in the light of our findings.

ACKNOWLEDGEMENTS

We thank Prof. Glenn Björk for advice and Prof. Göran Wadell for support. This study was financed by grants from the Swedish Medical Research Council (MFR), the Magnus Bergwall Trust and the Research Foundation of the Department of Oncology, University of Umeå.

REFERENCES

1 Yuh,C.-H., Chang,Y.-L. and Ting,L.-P. (1992) *J. Virol.*, **66**, 4073–4084.
2 Ganem,D., Pollack,J.R. and Tavis,J. (1994) *Infect. Agents Disease*, **3**, 85–93.
3 Bartenschlager,R., Junker-Niepmann,M. and Schaller,H. (1990) *J. Virol.*, **64**, 5324–5332.
4 Junker-Niepmann,M., Bartenschlager,R. and Schaller,H. (1990) *EMBO J.*, **9**, 3389–3396.
5 Hirsch,R.C., Lavine,J.E., Chang,L.J., Varmus,H.E. and Ganem,D. (1990) *Nature*, **344**, 552–555.
6 Hu,J. and Seeger,C. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 1060–1064.
7 Knaus,T. and Nassal,M. (1991) *Nucleic Acids Res.*, **21**, 3967–3975.
8 Pollack,J. and Ganem,D. (1993) *J. Virol.*, **67**, 3254–3263.
9 Laskus,T., Rakela,J. and Persing,D.H. (1994) *Virology*, **200**, 809–812.
10 Rieger,A. and Nassal,M. (1996) *J. Virol.*, **70**, 585–589.
11 Okamoto,H., Tsuda,F., Sakugawa,H., Sastrosoewignjo,R.I., Imai,M., Miyakawa,Y. and Mayumi, M. (1988) *J. Gen. Virol.*, **69**, 2575–2583.
12 Norder,H., Couroucé,A.-M. and Magnius,L.O. (1994) *Virology*, **198**, 489–503.
13 Naumann,H., Schaefer,S., Yoshida,C.F.T., Gaspar,A.M.C. and Gerlich, W.H. (1993) *J. Gen. Virol.*, **74**, 1627–1632.
14 Kidd-Ljunggren,K., Öberg,M. and Kidd,A.H. (1995) *J. Gen. Virol.*, **76**, 2119–2130.
15 Carman,W.F., Jacyna,M.R., Hadziyannis,S., Karayiannis,P., McGarvey,M.J., Makris,A. and Thomas,H.C. (1989) *Lancet*, **ii**, 588–590.
16 Okamoto,H., Tsuda,F., Akahane,Y., Sugai,Y., Yoshiba,M., Moriyama,K., Tanaka,T., Miyakawa,Y. and Mayumi,M. (1994) *J. Virol.*, **68**, 8102–8110.
17 Feitelson,M.A. (1994) *Lab. Invest.*, **71**, 324–349.
18 Horikita,M., Itoh,S., Yamamoto,K., Shibayama T., Tsuda,F. and Okamoto,H. (1994) *J. Med. Virol.*, **44**, 96–103.
19 Laskus,T., Rakela,J., Tong,M.J., Nowicki,M.J., Mosley,J.W. and Persing,D.H. (1994) *J. Hepatol.*, **20**, 837–841.
20 Lok,A.S.F., Akarca,U. and Greene,S. (1994) *Proc. Natl. Acad. Sci. USA*, **91**,4077–4081.
21 Li,J.-S., Tong,S.-P., Wen,Y.-M., Vitvitski,L., Zhang,Q. and Trepo,C. (1993) *J. Virol.*, **67**, 5402–5410.
22 Chen,I.-H., Huang,C.-J. and Ting,L.-P. (1995) *J. Virol.*, **69**, 3647–3657.
23 Kretz,K.A., Carson,G.S. and O′Brien,J.S. (1989) *Nucleic Acids Res.*, **17**, 5864.
24 Ljunggren,K. and Kidd,A.H. (1991) *J. Med. Virol.*, **34**, 179–183.
25 Higgins,G, Bleasby,A.J. and Fuchs,R. (1992) *Comput. Applicat. Biosci.*, **8**, 189–191.
26 Felsenstein,J. (1993) PHYLIP: phylogeny inference package V.3.5c. Distributed by the author, Department of Genetics, University of Seattle, Seattle, WA.
27 Genetics Computer Group Inc. (1995) Program manual of the Wisconsin Sequence Analysis Package V.8.1.
28 Jaeger,J.A., Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7706–7710.
29 Matzura,O. and Wennborg,A. (1995) RNADRAW V.1.1. http://mango.mef.ki.se/~ole/rnadraw/rnadraw.html.
30 Repp,R., Keller,C., Borkhardt,A., Csecke,A., Schaefer,S., Gerlich,W.H. and Lampert,F. (1992) *Arch. Virol.*, **125**, 299–304.
31 Barlet,V., Zarski,J.P., Thelu,M.A., Bichard,P. and Seigneurin,J.M. (1994) *J. Hepatol.*, **21**, 797–805.
32 Uchida,T., Gotoh,K. and Shikata,T. (1995) *J. Med. Virol.*, **45**, 247–252.
33 Gotoh,K., Mima,S., Uchida,T., Shikata,T., Yoshizawa,K., Irie,M. and Mizui,M. (1995) *J. Med. Virol.*, **46**, 201–206.
34 Tavis,J.E. and Ganem, D. (1995) *J. Virol.*, **69**, 4283–4291.
35 Russnak,R. and Ganem,D. (1990) *Genes Dev.*, **4**, 764–776.
36 Fiordialisi,G., Ghiotto,F., Castelnuovo,F., Primi,D. and Cariani,D. (1994) *J. Hepatol.*, **20**, 487–493.
37 Gerken,G., Kremsdorf, D., Capel.,F., Petit,M.A., Dauguet,C., Manns,M.P., Meyer zum Buschenfelde,K.-H. and Bréchot,C. (1991) *Virology*, **183**, 555–565.
38 Melegari,M., Bruno,S. and Wands,J.R. (1994) *Virology*, **199**, 292–300.
39 Nakajima,E., Minami,M., Ochiya,T., Kagawa,K., Okanoue,T. (1994) *J. Hepatol.*, **20**, 329–335.
40 Arii,M., Takada,S. and Koike,K. (1992) *Oncogene*, **7**, 397–403.
41 Takada,S., Kido,H., Fukutomi,A., Mori,T. and Koike,K. (1994) *Oncogene*, **9**, 341–348.
42 Takada,S. and Koike,K. (1994) *Virology*, 503–510.
43 Jacobs,B.L. and Langland,J.O. (1996) *Virology*, **219**, 339–349.